**Penerapan Metode K-Nearest Neighbor untuk Klasifikasi Kanker Paru**



**Niswatul Sifa**

**210411100145**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS TRUNOJOYO MADURA**

**2023**

**BAB 1**

**PENDAHULUAN**

* 1. **Latar Belakang**

Kanker paru-paru adalah salah satu kanker yang paling sering didiagnosis dan merupakan penyebab utama kematian di seluruh dunia [1]. Selain itu, angka tahun hidup disesuaikan dengan kecacatan (DALYs) tertinggi, dengan 40,9 juta orang pada tahun 2017. Periode kelangsungan hidup kanker adalah waktu yang berlalu antara diagnosis kanker dan kematian yang disebabkan oleh penyakit tersebut [2]. IARC (Badan Internasional untuk Penelitian Kanker) menjelaskan bahwa jumlah kasus baru kanker paru-paru di dunia pada tahun 2020 adalah 2,2 juta dan jumlah kematian baru akibat kanker paru-paru adalah 1,8 juta. Kelangsungan hidup lima tahun pasien kanker paru stadium awal dapat meningkat hingga lebih dari 90% setelah pengobatan, tetapi tingkat kelangsungan hidup pasien kanker paru stadium menengah adalah sekitar 60%, dan pasien kanker paru stadium lanjut bahkan kurang dari 5%. Akibatnya, diagnosis dini kanker paru sangat penting [3]. Kebiasaan merokok dalam jangka panjang dapat menyebabkan sebagian besar (85%) kasus kanker paru-paru, dan sekitar 10–15% kasus terjadi pada orang yang tidak pernah merokok [4]. Dibandingkan dengan jenis kanker paru-paru lainnya, wanita lebih rentan terhadap penyakit ini dibandingkan dengan pria, dan juga lebih sering menyerang orang yang lebih muda [5]. Kasus ini sering terjadi karena kombinasi faktor genetik dan paparan gas radon, asbes, perokok pasif, atau jenis polusi udara lainnya [4].

Para ilmuwan telah menerapkan berbagai teknik untuk mendiagnosis kanker paru-paru pada tahap awal salah satunya menggunakan biopsi. Tetapi, biopsi yang merupakan metode diagnostik tercanggih dan paling akurat juga dapat membahayakan pasien sehingga tidak dapat dilakukan secara berkala dalam waktu singkat [6]. Selain itu, ada juga pengobatan terapi seperti kemoterapi, radiasi, dan pembedahan dapat digunakan, meskipun hasilnya kurang memuaskan. Untuk meningkatkan efektivitas terapi dan mengurangi efek buruknya pada pasien, maka klasifikasi kanker paru-paru yang akurat sangat penting. Jutaan gen dapat diperiksa dengan teknologi microarray untuk mengetahui lebih banyak tentang cara sel bekerja. Data ini juga dapat digunakan untuk mendiagnosis dan memprediksi kanker [7]. Ada juga Computed Tomography (CT) yang merupakan metode non-invasif yang sering digunakan untuk deteksi dini, tetapi memiliki kelemahan, seperti radiasi pengion, biaya tinggi, dan hasil positif palsu. Oleh karena itu, sangat dibutuhkan metode diagnostik kanker paru-paru yang lebih murah, mudah digunakan, bebas radiasi, dan non-invasif [6].

Penelitian ini dilakukan untuk mengetahui apakah seorang pasien terdetesi penyakit kanker paru-paru Rendah, Sedang atau Tinggi. Adapun kumpulan data kanker paru-paru yang dipertimbangkan diperoleh dari situs Data world ( <https://data.world/cancerdatahp/lung-cancer-data> ) yang berisi 1000 sampel, masing-masing dengan 23 fitur. Untuk sasaran dalam dataset ini adalah tingkat risiko penderitaan kanker paru-paru yang diklasifikasikan dalam 3 tingkatan yaitu Rendah, Sedang, dan Tinggi apakah termasuk dalam salah satu kategori tersebut [4]. Selanjutnya akan dilakukan klasifikasi K-Nearest Neighbor (KNN) untuk mengklasifikasi jenis kanker paru-paru.

Untuk mengklasifikasi jenis kanker paru-paru apakah tergolong dalam kategori rendah, sedang, dan tinggi maka diperlukan suatu metode machine learning yaitu metode KNN. Karena algoritma pemilihan fitur dapat mempengaruhi kinerja model KNN, K-Nearest Neighbor (KNN) pertama kali dikembangkan untuk memprediksi kanker paru-paru pada tahap awal [4]. K-nearest-neighbor (KNN), yang merupakan metode nonparametrik sederhana sebagai dasar untuk klasifikasi, juga disukai ketika ada sedikit pengetahuan sebelumnya tentang data [8]. Selain itu, nilai parameter K pada algoritma KNN ditentukan secara eksperimental dengan menggunakan pendekatan iterative. Ketepatan metode KNN juga bisa berubah tergantung pada jumlah tetangga dan jumlah data yang digunakan untuk klasifikasi. KNN menunjukkan bahwa nilai parameter K yang berbeda dalam algoritma KNN dapat memengaruhi hasil algoritma secara signifikan. Untuk menemukan nilai K terbaik, pendekatan baru dikembangkan dalam lingkungan Python. Di bidang pengenalan pola, pengklasifikasi KNN telah digunakan secara luas. Pengklasifikasi tetangga terdekat bergantung pada pembelajaran melalui hubungan. Kedekatan pada algoritma KNN dapat ditandai dengan metrik pemisahan, misalnya jarak Euclidean [4]. K-Nearest Neighbor (KNN) bekerja karena jarak antara bidang pemisah dan titik data terdekat semakin kecil. [9].

Diantara penelitian-penelitian yang membahas tentang klasifikasi menggunakan metode KNN dengan topik yang berbeda-beda, saya akan membahas beberapa penelitian yang terkait dengan metode tersebut. Pada penelitian yang dilakukan oleh Ikhsan Wisnuadji Dkk yang membahas Tentang Implementasi Data Mining untuk Deteksi Penyakit Ginjal Kronis menggunakan K-Nearest Neighbor dengan Backward Elimination menghasilkan atribut yang telah terseleksi, yaitu berhasil menekan biaya pemeriksaan hingga 73,36%. Selanjutnya dilakukan pendeteksian penyakit menggunakan Algoritme K-Nearest Neighbors yang menghasilkan nilai akurasi sebesar 99,25%, sensitivity sebesar 99,5%, dan specificity sebesar 98,745% [10]. Pada penelitian lain yang dilakukan oleh Negar Maleki Dkk yang membahas tentang Metode K-NN Untuk Prognosis Kanker Paru-Paru Dengan Penggunaan Algoritma Genetika Untuk Pemilihan Fitur yaitu mendedikasikan 80% kumpulan data untuk pelatihan dan 20% untuk kumpulan pengujian. Eksperimen ini menghasilkan tingkat akurasi sebesar 100% untuk set pelatihan dan tingkat akurasi sebesar 96,2% untuk set pengujian. Hasil tersebut diperoleh dengan mengimplementasikan KNN dengan K sama dengan 10. Terlihat jelas bahwa perbedaan antara kedua tingkat akurasi ini tidak signifikan, oleh karena itu maka model tersebut dapat diterapkan. Selanjutnya, hasil menunjukkan bahwa pendekatan KNN tanpa menggunakan algoritma GA untuk memilih kombinasi fitur terbaik menghasilkan akurasi 96,2% ketika K-parameter diatur ke 10. Namun, akurasi ini lebih baik daripada yang diperoleh dengan metode pohon keputusan sebesar 95,2%, dan naik menjadi 99,8% ketika K diubah dari 10 menjadi 6 [4].

Dengan demikian, berdasarkan penjelasan yang telah disampaikan sebelumnya, maka penelitian ini bertujuan untuk mengklasifikasikan risiko pasien kanker paru-paru dalam tiga tingkatan yaitu rendah, sedang, dan tinggi menggunakan metode KNN dan juga untuk menentukan fitur terbaik yang meminimalkan kesalahan perhitungan pada metode KNN. Selain itu, terlihat bahwa ketika metode KNN digabungkan dengan algoritma pemilihan fitur, akurasi klasifikasi meningkat secara signifikan.

* 1. **Rumusan Masalah**

1. Bagaimana cara untuk mengaplikasikan Metode K-Nearest Neighbor (KNN) untuk mengklasifikasi risiko pasien kanker paru-paru dalam tiga tingkatan, yaitu rendah, sedang, dan tinggi?
2. Bagaimana pengaruh nilai parameter K dalam algoritma KNN terhadap akurasi klasifikasi kanker paru-paru?
   1. **Batasan Masalah**
3. Penelitian ini akan berfokus pada klasifikasi risiko pasien kanker paru-paru dalam tiga tingkatan, yaitu Rendah, Sedang, dan Tinggi. Pengklasifikasian tersebut akan dilakukan berdasarkan data yang tersedia.
4. Penelitian ini akan menggunakan metode K-Nearest Neighbor (KNN) untuk melakukan klasifikasi risiko kanker paru-paru. Aplikasi metode KNN akan digunakan untuk memprediksi tingkat risiko pasien.
5. Data yang digunakan dalam penelitian ini akan diperoleh dari situs Data world yang mencakup 1000 sampel dengan 23 fitur yang relevan untuk kanker paru-paru. Data ini akan digunakan sebagai basis untuk pelatihan dan pengujian model KNN.
   1. **Tujuan dan Manfaat**

Adapun tujuan dari penelitian ini adalah:

1. Untuk menerapkan metode KNN dalam klasifikasi risiko pasien kanker paru-paru agar menjadi tiga tingkatan, yaitu rendah, sedang, dan tinggi, dengan menggunakan data yang tersedia.
2. Untuk menganalisis pengaruh nilai parameter K terhadap akurasi klasifikasi kanker paru-paru menggunakan metode KNN.

Adapun manfaat dari penelitian ini adalah:

1. Dapat memberikan kontribusi dalam meningkatkan kemampuan diagnosis dini kanker paru-paru. Dengan menggunakan metode KNN dan fitur-fitur yang relevan, diharapkan pasien yang memiliki penyakit kanker paru dengan risiko tinggi dapat diidentifikasi lebih awal, dan juga agar memungkinkan pengobatan yang lebih efektif.
2. Dapat membantu mengurangi kesalahan diagnosis. Ini dapat membantu menghindari diagnosis palsu yang dapat berdampak pada pengobatan yang tidak sesuai.
   1. **Jadwal Penyelesaian**

Berikut merupakan timeline penyelesaian hingga berbentuk proposal:

|  |  |  |
| --- | --- | --- |
| **Timeline** | **Kegiatan** | **Penjelasan** |
| 30 September 2023 | Pencarian Data | Mencari data yang akan digunakan dalam penelitian |
| 1 – 4 Oktober 2023 | Study Literatur | Mencari jurnal dan review jurnal |
| 4 – 16 Oktober 2023 | Penyusunan Laporan | Menyusun Laporan Bab 1 |
| 17 – 24 Oktober 2023 | Penyusunan Laporan | Menyusun Laporan Bab 2 |
| 24 – 27 Oktober 2023 | Study Literatur | Mempelajari dan memahami proses perhitungan manual terkait dengan metode yang digunakan |
| 28 Oktober – 1 November 2023 | Perhitungan Manual | Melakukan perhitungan manual pada Cancer Patient Data Sets dengan metode KNN |
| 2 – 7 November 2023 | Pembuatan Diagram | Membuat Flowchart dan juga skenario pengujian |
| 7 – 14 November 2023 | Penyusunan Laporan | Menyusun Laporan Bab 3 |
| 14 – 30 November 2023 | Penyelesaian Proposal | Menyelesaikan laporan sampai berbentuk proposal yang baik dan benar |

**REFERENSI**

[1] Y. Li, X. Wu, P. Yang, Y. Luo, U. Illinois, and K. Kunci, “Pembelajaran Mesin untuk Diagnosis , Pengobatan , dan Prognosis Kanker Paru,” vol. 20, pp. 850–866, 2022, doi: 10.1016/j.gpb.2022.11.003.

[2] A. Sekolah, “Komputer dalam Biologi dan Kedokteran,” vol. 165, 2023, doi: 10.1016/j.compbiomed.2023.107338.

[3] L. H. A and G. H. B, “Heliyon,” vol. 9, pp. 1–14, 2023, doi: 10.1016/j.heliyon.2023.e13633.

[4] N. M. A, Y. Z. B, S. Taghi, A. Niaki, and A. Jurusan, “Sistem Pakar Dengan Aplikasi,” vol. 164, no. September 2020, pp. 1–7, 2021, doi: 10.1016/j.eswa.2020.113981.

[5] F. S. C and S. K. D. D, “Jurnal Informatika Patologi,” vol. 14, 2023, doi: 10.1016/j.jpi.2023.100307.

[6] A. Xianghaozhan, Z. W. A, M. Y. B, Z. Y. L. D, A. Wang, and G. L. A, “Pengukuran Prototipe diagnostik bantu berbasis hidung elektronik untuk deteksi kanker paru-paru dengan prediksi konformal,” vol. 158, 2020, doi: 10.1016/j.measurement.2020.107588.

[7] V. Yuvaraj and D. Maheswari, “Pengukuran : Sensor,” vol. 30, no. April, pp. 0–5, 2023, doi: 10.1016/j.measen.2023.100902.

[8] Y. X. A *et al.*, “Onkologi Translasi,” vol. 14, no. September 2020, 2021, doi: 10.1016/j.tranon.2020.100907.

[9] M. C. A, S. J. C. A, P. V. C, H. L. A, and E. O. A. A, “Seminar Biologi Kanker,” vol. 93, no. April, pp. 97–113, 2023, doi: 10.1016/j.semcancer.2023.05.004.

[10] I. W. Gamadarenda and I. Waspada, “Implementasi Data Mining untuk Deteksi Penyakit Ginjal Kronis (PGK) menggunakan K-Nearest Neighbor (KNN) dengan Backward Elimination,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 2, p. 417, 2020, doi: 10.25126/jtiik.2020721896.